B.Tech Even Semester Examination, May, 2025

Computer Science & Engineering

(8th Semester)

Course No: CSE-801(3)

(Deep Learning)

Full Marks: 50
Pass Marks: 25
Time: 2 hours

Note:

1. Attempt any five questions.
2. Begin each answer in a new page
3. Answer parts of a question at a place
4. Assume reasonable data wherever required
5. The figure in the right margin indicates full marks for the question
6. All the mathematical symbols and abbreviations have their usual meanings.

1. a) Which activation functions can lead to vanishing gradients? Write the equation of the activation function. (2)

   b) When should multi-task learning be used? (2)

   c) Explain why we need activation functions. (2)

   d) When implementing a neural network layer from scratch, we usually implement a 'forward' and a 'backward' function for each layer. Explain what these functions do, potential variables that they need to save, which arguments they take, and what they return (4)

2. (a) What is the primary purpose of regularization in a neural network? (2)

(b) Which regularization method leads to weight sparsity? Explain why. (3)

(c) Explain the concept of regularization in neural networks and compare L1 and L2regularization techniques . (5)

3. (a) Explain the concept of batch normalization. (2)

(b) Why is dropout used during the training of neural networks? (2)

(c) What is the purpose of early stopping in training neural networks? (2)

(d) Describe the saddle point problem in neural networks, its impact on training and strategies to mitigate it. (4)

4. (a) Explain the architecture and training process of a basic Autoencoder. (6)

(b) How autoencoders are used for denoising images? (4)

5. (a) Discuss the importance of weight initialization in training deep neural networks. Write a comparison of different weight initialization strategies. (6)

(b) You are designing a deep learning system to diagnose lung cancer through X-ray images. Which is the most appropriate evaluation metric and why: Accuracy, Precision, Recall, F1 score. (2)

(c) Why is it important to place non-linearity between the layers of neural networks? (2)

6. (a) What is a skip connection, and why is it used in neural networks? (3)

(b) What are the key components of a Transformer model in neural machine translation? (2)

(c) Compare and discuss grid search and random search for hyperparameter tuning in deep learning models (5)

7. (a) Given a Convolution Layer with 8 filters, a filter size of 6, a stride of 2, and a padding of 1. For an input feature map of 32 × 32 × 32, what is the output dimensionality after applying the Convolution Layer to the input? (2)

(b) What is the purpose of pooling? (2)

(c) Explain the different layers in CNN (6)

8. (a) Which type of neural network is primarily used in neural machine translation? (1)

(b) Explain how a bidirectional RNN works and what advantages it has over the unidirectional version (5)

(c) Explain what is the vanishing gradient problem in RNNs training and provide an example of one technique that aims at mitigate this problem (4)

*****